



# Eliminating Complexity to Ensure Fastest Time to Big Data Value

# Contents

Contents ..... 2

Introduction .....3

Problems Maintaining the Status Quo.....3

Pentaho as Your Technology Partner .....5

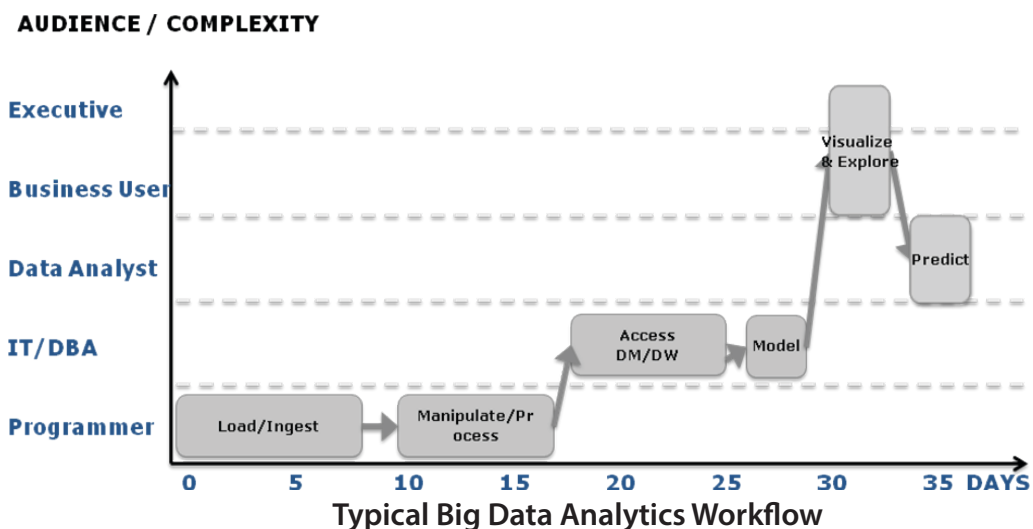
Contact Pentaho .....7

## Introduction

While most organizations recognize that big data analytics is key to business success, efforts are often stymied or slowed due to operational procedures and workflow issues. In fact, many companies find they now face the challenge of dealing with difficult, new and complex technologies.

In many cases, IT does not necessarily have enough people with the right specialized skills required to take on the process of turning big data into actionable, decision-making information. Often the need for these additional resources and skills are not anticipated.

Examining a typical big data analytics process workflow helps identify where many of these potential problems may occur, special skill sets are required and delays are introduced. Common steps in the big data analytics workflow include:



### LOADING AND INGESTION

Data must be obtained from a variety of both structured and unstructured data sources including documents, spreadsheets, email messages, images, texts, and social media content. While different organizations may take different approaches to loading and ingesting this data, most efforts require the development of custom code and the writing of scripts or use of specialized ETL tools to complete the process. These tasks are most often performed by developers or IT.

### MANIPULATION AND TRANSFORMATION

Once suitable data for the problem at hand is obtained, it must be prepared in the proper format before any analysis can be performed.

Even if this step requires something as straightforward as converting dollars to Euros, aggregating a set of data, or swapping rows of data for columns, the people carrying out the operations must have the expertise to complete the process. Manually completing these steps is a common, but time-consuming process that can easily introduce errors. Automating the processes typically requires the writing of some custom code.

### ACCESS

Once the data required for decision-making is in the correct format, it then needs to be directed to a big data store, to be accessed by business analytics applications.

Big data stores such as Hadoop, NoSQL, and analytic databases are the most common types and require specialized skills that many organizations do not currently have. This step in the workflow might also involve moving a slice of a larger dataset to a particular data warehouse. Again, special skills are needed to perform this step through coding or use of specialized ETL tools.

## MODEL

To derive actionable information from raw data, users need details about the database content. For example, when exploring issues related to inventories and sales, it is essential to know particular attributes (e.g., product ID, product name, price, etc.) of database entries and the relationships between the entities. A metadata model must be built that shows relationships and hierarchies to make the connection between data and business processes. This step might be done through custom coding or with a data modeling tool, but either approach requires expertise.

## VISUALIZE AND EXPLORE

The next step in the workflow is to examine relevant data, perform the actual analysis, and provide the information in the format needed for an executive, business unit head, or manager to make a quick and informed business decision. Typically users need different information and as they analyze data, requests for additional data and slices of data arise. Traditionally, developers or IT write code or use a BI tool to create fairly static views, dashboards, or reports based on the data.

## PREDICT

Once the data is visualized the next logical step in the big data work flow is to perform predictive analytics. Based on vast amounts of historical data predictive analytics can help organizations to better plan and optimize for the future.

## Problems With Maintaining the Status Quo

The current approach to big data analytics is rife with potential problems. The nature of each step in a common workflow requires manual intervention, opens up the process to potential mistakes, and delays the time to results. Worse, there is often a need for both IT and developers to do a great amount of hand coding and use specialized tools. Once the flow is completed and business users run analysis, new requests often come into IT for access to additional data sources and the linear process begins again. In today's budget-conscious and fast-paced business world, engaging these people can drain precious resources from other projects that are critical to the growth and success of the organization.

With companies dealing with an explosion of data volumes, while trying to incorporate more types of data into their decision-making processes, the problems, challenges, and pressure on resources simply multiply. At the same time, demand to make immediate decisions based on this information continues to increase. As a result, new thinking and new approaches are required.

Specifically, organizations need a solution that improves and shortens the big data analytics process workflow. The solution must have a number of characteristics to remove complexity, and simplify operations. It should also ensure that errors are not introduced and best practices are carried out.

To those points, a solution should offer easy to use data integration including support for structured and unstructured data. There should be tools for visualization and data exploration that address a broad set of users and can support multiple data sources so that both business and technical staff can quickly size up information and gauge its relevance and importance. The dynamic nature of today's marketplace means business priorities can quickly change.

A new opportunity may arise or a new data source (an organization's social media stream, for example) might provide needed insights into consumer interests. So the focus of big data analytics efforts will likely need to shift rapidly over time to keep pace with changing business opportunities. A solution that offers an easy to use visual development environment, rather than hand coding every time there is a change, would help an organization remain competitive.

And while it is a given that a big data analytics solution work with traditional data stores, a solution also must be capable of working with new big data stores such as Hadoop and NoSQL databases.

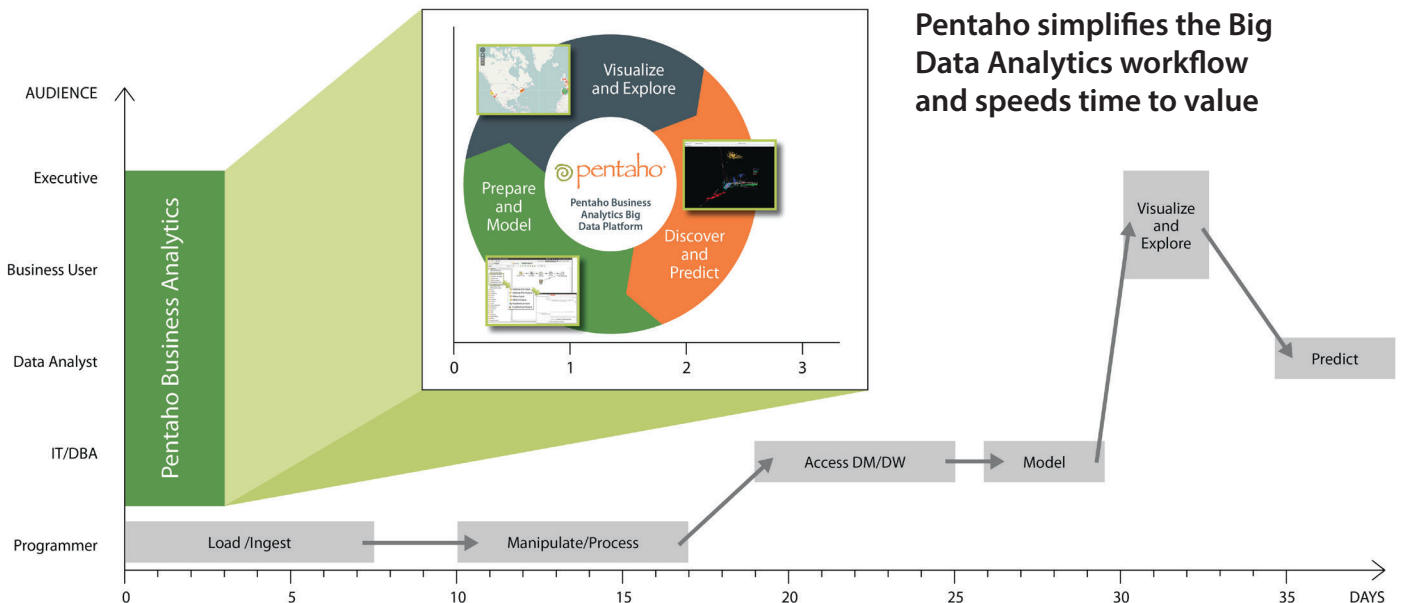
## Pentaho as Your Technology Partner

Clearly, big data analytics efforts need solutions that make processes easier to execute, do not require new technical resources, and reduce the time to results. This is where Pentaho can help. Pentaho offers a unified business analytics platform that supports the entire big data analytics flow.

In particular, Pentaho tightly couples data integration with business analytics in a single platform for both IT and business users. The Pentaho approach lets both groups easily access, integrate, visualize, and explore all data that impacts business results.

Pentaho Data Integration allows organizations to extract data from complex and heterogeneous sources and prepare that data for analysis all with a visual development environment that eliminates the need for specialized IT and developer resources. The solution produces consistent, high quality, ready-to-analyze data.

Additionally, Pentaho Business Analytics provides a highly interactive and easy to use web-based interface for business users to access and visualize data, create and interact with reports and dashboards, and analyze data, without depending on IT or developers. Beyond data discovery, Pentaho also offers predictive analytics capabilities as part of the platform.



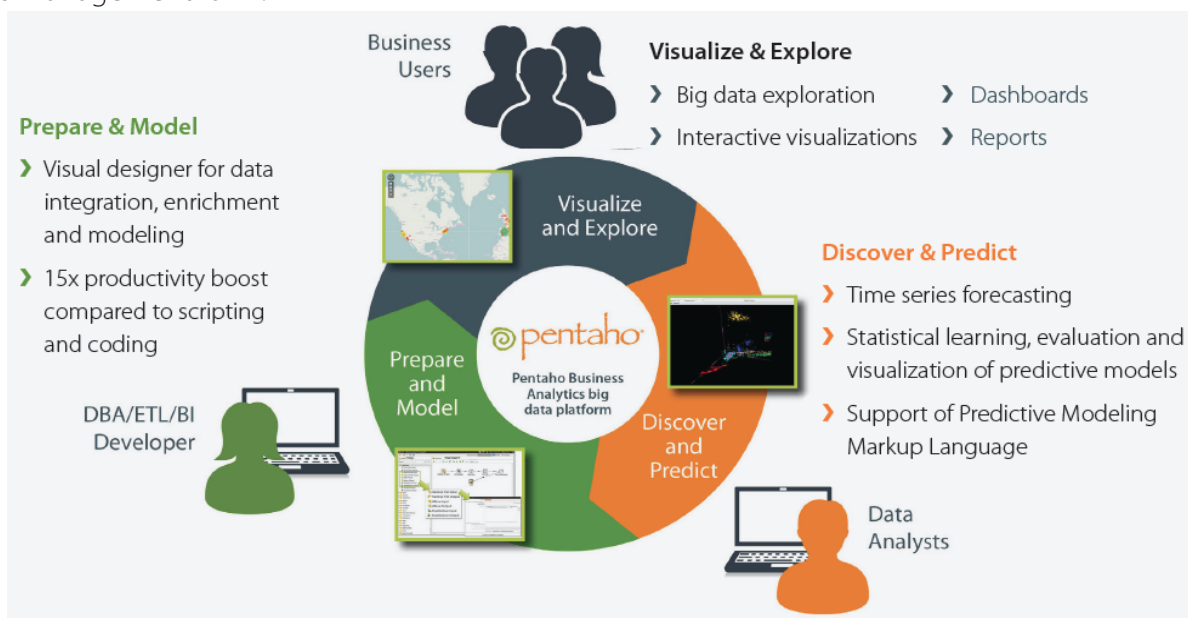
Thanks to those features, Pentaho software is currently being used by retailers, healthcare providers, media companies, hospitality organizations, and others. Clients include Beachmint, Shareable Ink, Social Commerce, TravelTainment, and Travian Games. (Access Pentaho's big data case studies and testimonials.)

One great differentiator that is getting the attention of such companies is Pentaho's visual development environment. Rather than writing code and developing scripts, organizations can quickly set up logical workflows to handle data ingestion, manipulation, integration, and modeling. This can greatly reduce the time to results, frees up staffing resources and opens up access to business analytics to many more users by removing technical complexity.

The most recent addition to the Pentaho platform is Instaview, the first instant and interactive application for big data. Instaview dramatically reduces the time and complexity required for data analysts to discover, visualize, and explore large volumes of diverse data in Hadoop, Cassandra, and HBase. Instaview broadens big data access to data analysts, removes the need for separate big data visualization tools, and simplifies big data delivery and access management for IT.

From an execution standpoint, Pentaho is unique in that it combines a visual development approach with the capability of being able to run in parallel as MapReduce across a Hadoop cluster. This results in executions that are as much as a 15 times faster versus using a hand-written code approach.

Given the potential problems that can crop up in managing and incorporating big data into decision-making processes, organizations need easy-to-use solutions that can address today's challenges, with the flexibility to adapt to meet future challenges. With a single platform that includes visual tools to simplify development, Pentaho not only shortens the time in getting to big data analytics, but also addresses the broadest set of users for business analytics capabilities such as data preparation, data discovery, and predictive analytics that support an iterative big data analytics process.



**Pentaho is the most complete solution for big data analytics**



#### **GLOBAL HEADQUARTERS**

Citadel International, Suite 340  
5950 Hazeltine National Dr.  
Orlando, FL 32822, USA

**TEL** +1 407 812 OPEN (6736)

**FAX** +1 407 517 4575

#### **US & WORLDWIDE SALES OFFICE**

201 Mission St., Suite 2375  
San Francisco, CA 94105, USA

**TEL** +1 415 525 5540

**TOLL FREE** +1 866 660 7555

#### **UNITED KINGDOM, REST OF EUROPE, MIDDLE EAST & AFRICA**

London, United Kingdom

**TEL** +44 7711 104854

**TOLL FREE** (UK) 0 800 680 0693

#### **France**

Offices - Paris, France

**TOLL FREE** (France) 0800 915343

#### **Germany, Austria, Switzerland**

Offices - Frankfurt, Germany

**TEL** +49 6051 7084 112

**TOLL FREE** (Germany) 0800 186 0332

#### **Belgium, Netherlands, and Luxembourg**

Offices - Antwerp, Belgium

**TEL** +31 621 505255

To learn more about Pentaho's Business Analytics software and services, contact your Pentaho Sales Representative online at **[pentahobigdata.com](http://pentahobigdata.com)** contact or call +1.866.660.7555.